

Directrices DRIVER para proveedores de contenidos

Proyecto Recolecta

25 de abril de 2008

Introducción

- Creación de servicios basados en los recursos alojados localmente.
- Orientadas a:
 - Recursos textuales: artículos científicos, tesis doctorales, ensayos, libros electrónicos y formatos similares para actividades de investigación científica.
 - Acceso abierto: acceso sin necesidad de ningún medio de pago, licencias, control de acceso con contraseña, control de acceso mediante pago, etc.

Introducción

Proveedor Driver validado. Recursos textuales

Obligatorio

- El repositorio contiene recursos textuales digitales
- Los recursos textuales están en formatos ampliamente utilizados y extendidos (PDF, TXT, RTF, DOC, TeX, etc.).
- Los recursos textuales son de acceso abierto, están disponibles directamente en el repositorio para cualquier usuario del mundo, y sin ninguna restricción (autorización, pago).
- Los recursos textuales están descritos mediante registros de metadatos.
- Los recursos textuales y de metadatos se vinculan entre sí de tal modo, que un usuario final pueda acceder al recurso textual mediante el identificador (normalmente una URL) del registro de metadatos.
- La URL de un recurso codificada en el registro de metadatos siempre se puede localizar y nunca se cambia ni se reasigna.
- Un identificador único identifica el registro de metadatos y el recurso textual (no hay punteros a sistemas externos, como un sistema bibliotecario nacional o un editor).

Introducción

Proveedor Driver validado. Recursos textuales

Recomendado

- Verificación transparente de la integridad de un recurso textual.
- Medidas de control de calidad (del contenido científico) de los recursos textuales expuestos para limitarlos a, por ejemplo, los recursos textuales incluidos en el informe científico anual (o equivalente).
- La URL de un recurso codificada en el registro de metadatos se basa en un esquema de identificadores persistentes (como DOI, URN, ARK, etc.).

Introducción

Proveedor Driver validado. Metadatos

Obligatorio

- Los metadatos se estructuran según la norma Unqualified Dublin Core (ISO 15836:2003).
- Los elementos individuales de DC se utilizan según lo dispuesto en el documento “Uso en DRIVER de Dublin Core”

.

Recomendado

- Los metadatos se estructuran según esquemas más completos, como Dublin Core Qualified o MODS.
- El idioma de los metadatos queda a la elección del proveedor de contenido. Se recomienda utilizar el inglés.
- El idioma recomendado para los resúmenes (los resúmenes son optativos) del artículo es el inglés.

Introducción

Proveedor Driver validado. OAI-MHP

Obligatorio

- El repositorio debe ser conforme con OAI y con la especificación “Uso en DRIVER de OAI-PMH”.
- Debe existir un identificador de repositorio y debe utilizarse el esquema de identificador OAI.
- Si (y sólo si) el repositorio contiene recursos que no sean obligatorios en el APARTADO A debe definirse un set OAI que identifique la colección de recursos textuales digitales con acceso directo.

Recomendado

- Previsiones de modificación de URL base .
- Respuesta completa a Identify, incluido el uso de la declaración Description .
- Uso del contenedor XML de DIDL.

Introducción

Anexos

- Directrices específicas para metadatos de DRIVER
- Directrices específicas para utilizar el OAI-PMH
- Directrices específicas para utilizar MPEG-21 DIDL (contenedor XML)

Anexo 1: Metadatos Dublin Core

- No son normas de catalogación.
- Directrices para la transformación de esquemas de metadatos a Dublin Core no cualificado (ISO 15836:2003).
- Codificación UNICODE.
- Recomendado el inglés (no obligatorio).
- Recomendado Dublin Core cualificado.
- Especificaciones para los 16 elementos DC:
 - Obligatorio > Obligatorio si procede > Recomendado > Optativo.

Anexo 1: Metadatos Dublin Core. Elemento Title

- Obligatorio.
- Nombre que se da al recurso. Normalmente el título es el nombre por el que el recurso es conocido normalmente.
- Separar subtítulos mediante dos puntos.
- Mayúsculas solo para nombres propios.

En Recolecta:

- 99,99% de los registros contemplan la etiqueta "Title".
- Uso de mayúsculas/minúsculas arbitrario, ...

Anexo 1: Metadatos Dublin Core. Elemento Creator

- Obligatorio.
- La entidad principal responsable de crear el contenido del recurso.
- Apellidos, Nombre, Iniciales.
- Jerarquía en las organizaciones: de mayor a menor separada por puntos.

En Recolecta:

- 92,49% de los registros contemplan la etiqueta “Creator”.
- Formas diversas (apellidos, nombre; nombre apellidos, etc...).

Anexo 1: Metadatos Dublin Core. Elemento Subject

- Obligatorio si procede.
- La materia del recurso.
- Palabra clave o clasificación.
- Evitar palabras generales para la definición de materias.
- Separar palabras clave por “;” o repetir el campo.
- No confundir con “Type”.

En Recolecta:

- 29,99% de los registros contemplan la etiqueta “Subject”.

Anexo 1: Metadatos Dublin Core. Elemento Description

- Obligatorio si procede.
- Información del contenido del recurso.
- Puede contener: resumen, índice de contenido, referencias a representaciones gráficas del contenido o texto libre, ...

En Recolecta:

- 41,47% de los registros contemplan la etiqueta "Description".
- En algunos casos incorporación de marcado HTML en las descripciones.

Anexo 1: Metadatos Dublin Core. Elemento Publisher

- Obligatorio si procede.
- Entidad responsable de hacer que el recurso esté disponible (persona, organización, servicio).
- No confundir con: afiliación del autor, “Contributor” o “Creator”.
- Jerarquía en las organizaciones: de mayor a menor separada por puntos.

En Recolecta:

- 91,31% de los registros contemplan la etiqueta “Publisher”.

Anexo 1: Metadatos Dublin Core. Elemento Contributor

- Optativo.
- Entidad responsable de contribuir al contenido del recurso (persona, organización, servicio).
- Por ejemplo: directores de tesis, tutores, ...
- A nivel forma: ver “Creator”.
- No confundir con “Creator” o “Publisher”.

En Recolecta:

- 6,25% de los registros contemplan la etiqueta “Contributor”.

Anexo 1: Metadatos Dublin Core. Elemento Date

- Obligatorio.
- Fecha asociada al ciclo de vida del recurso.
Normalmente se asocia a la creación o la disponibilidad del recurso.
- Formato: AAAA-MM-DD.
- Utilizar la etiqueta “Date” para la fecha de creación.

En Recolecta:

- 100% de los registros contemplan la etiqueta “Date”.
- Distintas fechas no identificados (creación, publicación, disponibilidad,...).

Anexo 1: Metadatos Dublin Core. Elemento Type

- Obligatorio.
- Tipo de resultado científico del cual el recurso es una manifestación.
- Se contemplan 14 tipos distintos de documentos.
- No confundir con “Format”.

En Recolecta:

- 93.39% de los registros contemplan la etiqueta “Type”.
- 145 tipos distintos utilizados.

Anexo 1: Metadatos Dublin Core. Elemento Type

Type DRIVER	Tipo Supuesto	Con Type Driver	%
Article	246639	8121	3,29%
Conference lecture	661	36	5,45%
Conference report	234	0	0,00%
Contribution for newspaper or weekly	0	0	
Doctoral thesis	6751	155	2,30%
Master thesis	21	21	100,00%
Bachelor thesis	3358	1767	52,62%
External research report	2151	22	1,02%
Lecture	361	0	0,00%
Internal report	173	71	41,04%
Newsletter	2	0	0,00%
Part of book or chapter of book	649	87	13,41%
Research paper	2246	707	31,48%

- Tipos no contemplados (cartas, datasets, revistas, revisado por pares, ...)
- Poca presencia de materiales docentes.
- Uso de "Type" para indicar formatos.

Anexo 1: Metadatos Dublin Core. Elemento Format

- Recomendado.
- Manifestación física o digital del recurso (medio, dimensión, ...).
- Tipos MIME (<http://www.iana.org/assignments/media-types/>).
- No confundir con “Type”.

En Recolecta:

- 41,61 % de los registros contemplan la etiqueta “Format”.
- 46,45 % para indicar tipos MIME.
- Otros usos: tamaño de archivos, número de páginas, ...

Anexo 1: Metadatos Dublin Core. Elemento Identifier

- Obligatorio.
- Referencia inequívoca al recurso.
- URI, URL, DOI,...
- No confundir con “Source” o “Relation”.
- En Recolecta:

En Recolecta:

- 100% de los registros contemplan la etiqueta “Identifier”.
- Uso habitual de URI/URL, ISSN para artículos, ...

Anexo 1: Metadatos Dublin Core. Elemento Source

- Optativo.
- Referencia al recurso del que se deriva el recurso actual.
- Recomendado: utilizar solo si el recurso es resultado de la digitalización de originales no digitales.
- No confundir con “Identifier” o “Relation”.

En Recolecta:

- 84,14% de los registros contemplan la etiqueta “Source”.
- Referencia a números de revistas, ...

Anexo 1: Metadatos Dublin Core. Elemento Language

- Recomendado.
- Idiomas del contenido intelectual del recurso.
- ISO 639-3.
- Repetir si se encuentra en varios idiomas.

En Recolecta:

- 93,3% de los registros contemplan la etiqueta "Language".
- 82,71% según ISO 639-3 y no repetido.

Anexo 1: Metadatos Dublin Core. Elemento Relation

- Optativo.
- Referencia a un recurso relacionado.
- Identificar el recurso relacionado por códigos formales.

En Recolecta:

- 5,92% de los registros contemplan la etiqueta “Relation”.
- Uso similar a “Source”. Se usan indistintamente.
- Enlace a los archivos del documento.

Anexo 1: Metadatos Dublin Core. Elemento Coverage

- Optativo.
- Alcance o ámbito del contenido del recurso (ubicación espacial, tiempo, jurisdicción,...).
- Preferentemente uso de vocabularios controlados.
- Repetir para hacer referencia a múltiples ámbitos.

En Recolecta:

- 0,44% de los registros contemplan la etiqueta "Coverage".
- Códigos de países o fechas.

Anexo 1: Metadatos Dublin Core. Elemento Rights

- Recomendado.
- Información sobre los derechos contenidos en y sobre los recursos.

En Recolecta:

- 86,50% de los registros contemplan la etiqueta "Rights".
- Uso de fórmulas diversas.

Anexo 1: Metadatos Dublin Core. Elemento Audience

- Optativo.
- Clase de entidad para la que el recurso está diseñado o puede ser de utilidad.

En Recolecta:

- 0,7% de los registros contemplan la etiqueta "Audience".
- Uso para recursos educativos, ...

Anexo 1: Metadatos Dublin Core. Elemento Tabla resumen

Etiqueta D.C	No usado	Usado	Tipo	% uso
Title	17	295652	Obligatorio	99,99%
Creator	22218	273451	Obligatorio	92,49%
Subject	207010	88659	Obligatorio si procede	29,99%
Description	173066	122603	Obligatorio si procede	41,47%
Publisher	25685	269984	Obligatorio si procede	91,31%
Contributor	277189	18480	Optativo	6,25%
Date	0	295669	Obligatorio	100,00%
Type	19555	276114	Obligatorio	93,39%
Format	172632	123037	Recomendado	41,61%
Identifier	6	295663	Obligatorio	100,00%
Source	46900	248769	Optativo	84,14%
Language	19822	275847	Recomendado	93,30%
Relation	278176	17493	Optativo	5,92%
Coverage	294367	1302	Optativo	0,44%
Rights	39914	255755	Recomendado	86,50%
Audience	293596	2073	Optativo	0,70%

Amplio uso de las etiquetas obligatorias

Anexo 2: Protocolo OAI-MHP

- Seguir la versión 2.0 del protocolo.
- Registros eliminados: se deben declarar. Utilizar la opción “transient” o “persistent” (conservar temporalmente o indefinidamente la información sobre el registro eliminado).
- ResumptionToken (testigos de reanudación) al menos conservados durante 24 h.

Anexo 2: Protocolo OAI-MHP

- Lotes de recolección entre 100 y 200 registros.
- Correo electrónico de administración disponible.
- Comunicar cambios en la URL base, sets, formatos de metadatos, ...

Anexo 2: Protocolo OAI-MHP

- Set Driver
 - Solo si es necesario.
 - Agrupa los registros textuales y de acceso abierto.
 - setName: Open Access DRIVERset; setSpec: driver

Anexo 3: DIDL

- Origen: DARE. Implementado en Holanda desde 2006.
- Facilita el transporte de los objetos digitales que forman parte de un ítem de un repositorio.
- Contenedor recolectable por OAI-MHP.
- Útil para la incorporación de contenidos digitales a otros sistemas y para el desarrollo de servicios avanzados (búsqueda al texto completo, análisis de documentos para la extracción de citas, ...).

Referencias

- DRIVER Project (2007). [Directrices del proyecto Driver para los proveedores de contenido, anexos y consejos de implementación](http://www.recolecta.net/buscador/documentos.jsp)
<<http://www.recolecta.net/buscador/documentos.jsp>>
- DINI (2007). [Traducción certificado de DINI](http://www.recolecta.net/buscador/documentos.jsp). Servicio de Documentación y Publicación
_<<http://www.recolecta.net/buscador/documentos.jsp>>
- DRIVER : Digital Repository Infrastructure for European Research
<<http://www.driver-community.eu/>>